平均型基準の多目的マルコフ決定過程について(学習と制御とその周辺)

AUTHOR(S):

伊喜, 哲一郎

CITATION:

伊喜, 哲一郎. 平均型基準の多目的マルコフ決定過程について(学習と制御とその周辺). 数理解析研究所講究録 1985, 557: 14-31

ISSUE DATE:

1985-04

URL:

http://hdl.handle.net/2433/98988

数理解析研究所講究録
557巻 1985年 14-31

14

平均型基準の多目的マルコフ決定過程について

宮崎大教育　伊喜　哲一郎（Tetsuichiro IKI）

利得が多目的でベクトル値として与えられるマルコフ決定過程について述べている。直接期待利得の平均型基準値に対してパレート最適政策を求めている。系は有限個の状態間を推移し，各状態では有限個の決定を行う。パレート錐よりも，もっと一般の凸錐を用いた最適方程式が導入されている。拡張されたハワードの政策改良法により最適方程式の解集合を定める。凸最適政策はこの解集合の最大要素を与える定常政策として，またパレート最適政策はその特別の場合のものとして与えられている。

§1. はじめに

有限個の状態間を推移し，各状態で許される決定の数も有限個である定常的マルコフ決定過程について述べる。直接期待利得は多目的でベクトル値として与えられるとする。

後に定義を述べる時間平均利得を平均型基準値と呼び，そのパレート最適政策を求める。従来の平均型基準マルコフ決定過程に関する研究は，単一目的の場合についてであった。多目的で割引因子を持つ問題に対しては，N. Furukawa [3],[4] の研究がある。平均型基準値の問題に対して，L.C. Thomas [6] は，全ての政策の下で系が unichain をなす場合にパレート最適政策を与えている。T. Iki & N. Furukawa [5] では，系が多重連鎖をなす場合に与えている。本報告は[5]の結果を拡張してパレート錐よりもっと一般の凸錐を用いた場合について述べている。凸錐の拡張により，与えられたベクトル値を凸錐の一つである半空間によってスカラー化し，目的とするパレート最適解の第1近似解集合を求め，これを基礎として直接パレート解を求めることが可能となる。最適方程式は[5]におけるものよりも一般化される。ハワードの政策改良法により最適方程式の解集合を定める。拡張された意味の凸最適政策はこの解集合の最大要素を与える定常政策として求まる。またパレート最適政策は，凸錐をパレート錐に限定した特別の場合として与えられる。

§2で問題の定式化を，§3で政策更改の手続きを，§4で政策改良法の理論付を行ったのち，§5で結論を述べ，§6でアルゴリズムを提示する。

§2. 問題の定式化

p を自然数，$\mathbb{R}^p$ を p 次元ユークリッド空間とする。多目的マルコフ決定過程を $\{S, A, (q_{ij}^a), (r_j^a)\}$ で表わす。ただし，各記号は次の通りとする。$S \equiv \{i \mid i=1,2,\cdots,N\}$ は N 個 $(N \geq 1)$ の要素から成る状態空間。A_i は状態 i でとりうる決定の集合とし，A は各 A_i の直積空間であり $A = \underset{i \in S}{\times} A_i$。$q_{ij}^a$ は状態 i で決定 a をとった時に系が状態 j へ1歩で推移する確率。(q_{ij}^a) は推移確率行列。r_j^a は状態 j で決定 a をとった時の直接期待利得であり，空間 $\mathbb{R}^p$ における要素。(r_j^a) は N 状態 p 重ベクトル。

写像 $f : S \longmapsto A$ に対して，列 $(f, f, \cdots)$ を定常政策と呼び再び便宜的に f で表わす。定常政策全体の集合を F で表わす。定常政策を以後は単に政策と呼ぶことにする。各 $f \in F$ に対して記号 $Q(f)$，$r(f)$ を $q_{ij}^f \equiv q_{ij}^{f(i)}$，$Q(f) \equiv (q_{ij}^f)$，$r(f)_i \equiv r_i^{f(i)}$，$r(f) \equiv (r(f)_i)$ によって導入する。

X を有限集合とするとき，$M^p(X)$ によって X 上で定義された有界な p 重ベクトル値関数の全体から成る集合を表わす。

仮定：　各 $f \in F$ について　$r(f) \in M^p(S \times F)$ である

ことを仮定する。各 $x = (x_1, x_2, \cdots, x_p) \in \mathbb{R}^p$ について，全ての $\ell (= 1, 2, \cdots, p)$ に対して $x_\ell \geq 0$ であるとき $x \geqq 0$ と表わす。0 は $\mathbb{R}^p$ の原点とする。$\mathbb{R}_+^p \equiv \{x \in \mathbb{R}^p \mid x \geqq 0\}$ とおく。

$\mathbb{R}^p_+$をパレート錐と呼ぶ。mを自然数とする。$m \times p$次の行列Wを与える。$K \equiv \{x \in \mathbb{R}^p \mid Wx \geqq 0_m\}$と定義する。ただし$0_m$は$\mathbb{R}^m$の原点であるが，以後原点の次元は一々明記しない。$K$は$\mathbb{R}^p$の原点を頂点とする凸多面錐である。行列$W$が$p \times p$次の単位行列ならば$K = \mathbb{R}^p_+$であり，とくに$m=1$の時には，$K \supset \mathbb{R}^p_+$である開半空間をなすように$W$を選ぶことができる。$\Lambda_0 \equiv K \cap (-K)$, $\Lambda_1 \equiv K \cap (-K)^c$, $\Lambda \equiv \Lambda_1 \cup \{0\}$とおくと次の補題を得る。

補題1.（Y.C. Wong & K.F. Ng [8]）

(i) $0 \in \Lambda_0$，(ii) $0 \notin \Lambda_1$かつΛ_1は凸集合である。

(iii) $\Lambda \cap (-\Lambda) = \{0\}$かつ$\Lambda$は$\Lambda_1$を含む最小の凸錐である。

Kは最適性の判定，Λは最適方程式の導入，Λ_1は政策改良法において有用である。

各$f \in F$に対して，$Q(f)$のチェザロ極限行列$Q^*(f)$は$Q^*(f) \equiv \lim_{n\to\infty} (n+1)^{-1} \sum_{k=1}^{n} Q^k(f)$によって定義される。その第$i$-$j$成分を$\pi^f_{ij}$と表わす。$r(f)$の時間平均は$u(f) \equiv Q^*(f) r(f)$と定義される。$u(f)$を平均型基準値と呼ぶことにする。また基本行列$H(f)$は$H(f) \equiv [I - Q(f) + Q^*(f)]^{-1} - Q^*(f)$によって，$u(f)$の相対値は$v(f) \equiv H(f) r(f)$と定義される。$u(f)$, $v(f)$の各々の第i成分を$u(f)_i$, $v(f)_i$で表わす。

空間R^pの要素x, yに対して，$x-y\in K$であるときに限り$x\geqq y$，また$x-y\in\Lambda_1$であるときに限り$x\geq y$と記す。対称的に$\leqq$，$\leq$を導入する。$x-y\in\Lambda_0$であるときに限り$x\simeq y$と記す。ここで K－最適政策の概念を導入する。

定義1. $f^*\in F$が関係

$$(\forall f\in F)(\forall i\in S)(u(f)_i\geqq u(f^*)_i\longrightarrow u(f)_i\simeq u(f^*)_i)$$

を満足するとき，K－最適政策と呼ぶ。

$K=R^p_+$のときはパレート最適であることを，原点0を境界点に持つ閉半空間のときはスカラー化最適であることを意味する。平均型基準値に対してのみ最適性の判定を行い，相対値の優劣を直接的に評価することなく，政策改良法により全てのK－最適政策を求める方法を述べるのが，本報告の主目的である。

§3. 基本補題と政策更改手続き

従来の単一目的に対するマルコフ決定過程論の中から，必要となる基本的結果を列記する。

補題2 (D. Blackwell [1], A.F. Veinott, Jr. [7])

各$f\in F$に対して，$u(f)$は連立一次方程式

$$[I-Q(f)]u=0,\quad Q^*(f)u=Q^*(f)r(f)$$

の唯一解であり，$v(f)$は連立一次方程式

$$[I-Q(f)]v=r(f)-u(f),\quad Q^*(f)v=0$$

の唯一解である。

各 $f\in F$ に対して各々の第1方程式から $u(f)=Q(f)u(f)$ $r(f)+Q(f)v(f)=u(f)+v(f)$ が成立することを導く。これを基に，各 f, $g\in F$ に対して $\Delta(g,f)\equiv Q(g)u(f)-u(f)$, $L(g,f)\equiv r(g)+Q(g)v(f)-u(f)-v(f)$ とおくと，次の補題および系を得る。

補題3.（D. Blackwell [1], A.F. Veinott, Jr. [7]）

任意の f, $g\in F$ に対して

(i) $$u(g)-u(f)=\Delta(g,f)+Q(g)[u(g)-u(f)]$$

(ii) $$u(g)-u(f)+v(g)-v(f)=L(g,f)+Q(g)[v(g)-v(f)]$$

が成立する。

系3.1. 任意の $f,g\in F$ に対して

(i) $$Q^*(g)\Delta(g,f)=0$$

(ii) $$Q^*(g)[u(g)-u(f)]=Q^*(g)L(g,f)$$

が成立する。

任意の $f\in F$ と各 $i\in S$ および $a\in A_i$ に対して

$$\Delta(a,f)_i\equiv\sum_{j\in S}q_{ij}^a u(f)_j-u(f)_i$$

$$L(a,f)_i\equiv r_i^a+\sum_{j\in S}q_{ij}^a v(f)_j-u(f)_i-v(f)_i$$

と定義する。$a=g(i)$ ならば $\Delta(a,f)_i$ は $\Delta(g,f)$ の第 i 成分，$L(a,f)_i$ は $L(g,f)$ の第 i 成分である。特に $g(i)=f(i)$ な

らば $\Delta(g,f)_i=0$, $L(g,f)_i=0$ である。 各 $i\in S$ に対して以下の集合を導入する。 $G_1(i,f)\equiv\{a\in A_i \mid \Delta(a,f)_i \geq 0\}$, $G_2(i,f)\equiv\{a\in A_i \mid \Delta(a,f)_i=0$ かつ $L(a,f)_i \geq 0\}$, $G(i,f)\equiv G_1(i,f)\cup G_2(i,f)$, 対称的に $D_1(i,f)\equiv\{a\in A_i \mid \Delta(a,f)_i \leq 0\}$, $D_2(i,f)\equiv\{a\in A_i \mid \Delta(a,f)_i=0$ かつ $L(a,f)_i \leq 0\}$, $D(i,f)\equiv D_1(i,f)\cup D_2(i,f)$ および $B(i,f)\equiv G(i,f)^c\cap D(i,f)^c$ と定義する。さらに各 $f\in F$ に対して $S_1(f)\equiv\{i\in S \mid G_1(i,f)\neq\emptyset\}$, $S_2(f)\equiv\{i\in S \mid G_2(i,f)\neq\emptyset\}$, $S_0(f)\equiv S_1(f)\cup S_2(f)$, 対称的に $\tilde{S}_1(f)\equiv\{i\in S \mid D_1(i,f)\neq\emptyset\}$, $\tilde{S}_2(f)\equiv\{i\in S \mid D_2(i,f)\neq\emptyset\}$ および $\tilde{S}_0(f)\equiv\tilde{S}_1(f)\cup\tilde{S}_2(f)$ と定義する。

順序 $\geq$, $\leq$ を用いた政策更改の手続きを次に定義する。

定義2. 各 $f\in F$ に対し, もし $S_0(f)\neq\emptyset$ ならば, 各 $i\in S_0(f)$ において任意の $a\in G(i,f)$ をとり $g(i)=a$; 各 $i\in S-S_0(f)$ に対しては $g(i)=f(i)$ とする。この様な g 全体の集合を $G(f)$ と表わす。(政策改良の手続き)

定義3 各 $f\in F$ に対し, もし $\tilde{S}_0(f)\neq\emptyset$ ならば, $i_0\in\tilde{S}_0(f)$ においては任意の $a\in D(i_0,f)$ をとり $d(i_0)=a$; $i\in\tilde{S}_0(f)-\{i_0\}$ においては任意の $a'\in D(i,f)\cup\{f(i)\}$ をとり $d(i_0)=a'$; その他の $i\in S-\tilde{S}_0(f)$ 対しては $d(i)=f(i)$ とする。この様な d 全体の集合を $D(f)$ と表わす。

集合 $D(f)$ は f の回りで非最適政策として除去可能性の判定を

行う対象となる政策の集合である。

定義4. 各$f\in F$に対し，政策$b\in F$は，もし$b\neq f$，ある$i_0\in S$において$b(i_0)\in B(i_0,f)\cap\{f(i_0)\}^c$および各$i\in S-\{i_0\}$において$b(i)\in B(i,f)$となっているならば，$f$の回りのdoubtful policyと呼ばれる。この様なb全体の集合を$B(f)$と表わす。

もし政策改良が$f\in F$で停止されたならば$B(f)$を必要とする。

§4. 主要定理

本節では任意の$f\in F$に対し，§3における$G(f)$と$D(f)$を対象として凸集合Λ_1を用いた政策改良法について述べる。以下では任意に$f\in F$をとり固定しておく。

再帰的状態の集合と過渡的状態の集合を$C_e(f)\equiv\{i\in S\mid\pi_{ii}^f>0\}$，$C_0(f)\equiv\{i\in S\mid\pi_{ii}^f=0\}$によって表わす。$C_0(f)$に対応した推移確率行列を$Q_0(f)$と表わすと逆行列$[I-Q_0(f)]^{-1}$が存在することは良く知られているので，その第$i$-$j$成分を$n_{ij}^f$と表わす。記号$t_{ij}^f$を，任意の$(i,j)\in C_0(f)\times C_0(f)$に対しては$t_{ij}^f\equiv n_{ij}^f$，その他の$(i,j)\in[S\times S-C_0(f)\times C_0(f)]$に対しては$t_{ij}^f\equiv 0$と定義する。行列$(t_{ij}^f)$を$T(f)$と表わす。有限状態のマルコフ決定過程ではつねに$C_e(f)\neq\emptyset$である。またもし$C_0(f)\neq\emptyset$ならば，$i\in C_0(f)$に対しては$n_{ii}^f>1$，$(i,j)\in C_0(f)\times C_0(f)$に対しては$n_{ij}^f\geq 0$で

ある。任意に $g\in G(f)$ をとり f に対すると同様に $C_0(g)$, $Ce(g)$, $T(g)$ および $Q^*(g)$ を作ると次の補題を得る。

補題4. もし $S_1(f)\neq\emptyset$ であるか，又はもし $S_1(f)=\emptyset$ かつ

$S_2(f)\cap Ce(g)\neq\emptyset$ であるならば

(i) $$S_1(f)\subset C_0(g)$$

(ii) $$u(g)-u(f)=T(g)\Delta(g,f)+Q^*(g)L(g,f)$$

が成立する。とくに $S_1(f)\neq\emptyset$ ならば，各 $i\in S_1(f)$ に対し

(iii) $$u(g)_i-u(f)_i=\sum_{j\in S_1(f)}n^g_{ij}\Delta(g,f)_j+\sum_{j\in S_2(f)\cap Ce(g)}\pi^g_{ij}L(g,f)_j\geq 0$$

が成立する。また $S_1(f)=\emptyset$ かつ

$S_2(f)\cap Ce(g)\neq\emptyset$ ならば各 $i\in S_2(f)\cap Ce(g)$ に対し

(iv) $$u(g)_i-u(f)_i=\sum_{j\in S_2(f)\cap Ce(g)}\pi^g_{ij}L(g,f)_j\geq 0$$

が成立する。

(証明) (i)を示すために任意に $k\in S_1(f)$ をとる。もし $\pi^g_{kk}>0$ とすると $\pi^g_{kk}\Delta(g,f)_k\in\Lambda_1\subset\Lambda$ である。一方系3.1の(i)より $\pi^g_{kk}\Delta(g,f)_k=-\sum_{j\in S_1(f)-\{k\}}\pi^g_{kj}\Delta(g,f)_j$。よって $\pi^g_{kk}\Delta(g,f)_k\in\Lambda\cap(-\Lambda)$ となり矛盾。ゆえに $\pi^g_{kk}=0$。こうして $S_1(f)\subset C_0(g)$ を知る。

(ii)を示すために補題3の(i)をくり返し用いると

$$u(g)-u(f)=\sum_{n=0}^{m}Q^n(g)\Delta(g,f)-\frac{1}{m+1}\sum_{n=1}^{m}nQ^n(g)\Delta(g,f)+\frac{1}{m+1}\sum_{n=1}^{m+1}Q^n(g)[u(g)-u(f)]$$

と変形できるので上の(i)および系3.1の(iii)より $m\to\infty$ とすると(ii)を得る。(iii), (iv)は(ii)より明らかである。

補題5. $S_1(f)=\emptyset$, $S_2(f)\neq\emptyset$ かつ $S_2(f)\cap Ce(g)=\emptyset$ ならば

(i) $u(g)=u(f)$

(ii) $Ce(g)\subset S_0^c(f)$ かつ 各 $(i,j)\in Ce(g)\times S$ に対し $\pi_{ij}^g=\pi_{ij}^f$

(iii) $v(g)-v(f)=T(g)L(g,f)$

が成立する。とくに各 $i\in S_2(f)$ に対しては

(iv) $v(g)_i-v(f)_i=\sum_{j\in S_2(f)} n_{ij}^g L(g,f)_j \geq 0$

が成立する。

(証明) 参考文献 [5] と同様に示すことができる。

ここで $M^p(S)$ 上に順序 $\underset{M}{\leqq}$, $\underset{M}{<}$ を導入する。$M^p(S)$ における任意の要素を φ, ψ とし, $\varphi=(\varphi_1,\varphi_2,\cdots,\varphi_N)$, $\psi=(\psi_1,\psi_2,\cdots,\psi_N)$ とすると各 $i\in S$ に対し $\varphi_i\in R^p$, $\psi_i\in R^p$ である。もし全ての $i\in S$ において $\varphi_i\leqq\psi_i$ ならば $\varphi\underset{M}{\leqq}\psi$ と表わす。もし $\varphi\underset{M}{\leqq}\psi$ かつ $\varphi_{i_0}\leq\psi_{i_0}$ なる $i_0\in S$ が少くとも1つ存在するならば $\varphi\underset{M}{<}\psi$ と表わす。さらに $M^p(S)\times M^p(S)$ 上に辞書式順序 $\lneqq$ を導入する。任意の対 $(u',v')\in M^p(S)\times M^p(S)$ と $(u^2,v^2)\in M^p(S)\times M^p(S)$ に対して, もし $u'\underset{M}{<}u^2$ であるかまたは $u'=u^2$ かつ $v'\underset{M}{<}v^2$ ならば $(u',v')\lneqq(u^2,v^2)$ と記す。

この時, 補題4. と5. より次の主要定理を得る。

定理1. 任意の $f\in F$ に対し

(i) もし $S_0(f)\neq\emptyset$ ならば, 任意の $g\in G(f)$ に対して

$$(u(f),v(f))\lneqq(u(g),v(g)).$$

(ii) もし $\hat{S}_0(f) \neq \emptyset$ ならば，任意の $d \in D(f)$ に対して

$$(u(d), v(d)) \leqq (u(f), v(f)).$$

(証明) (i)は直接的であり，(ii)は順序 $\leqq$ を対称的に用いる。

この定理は単一目的の平均型基準値問題に対するハワードの政策改良法に関する結果を，ベクトル値問題において R_+^l よりももっと広い錐 K へと拡張した結果を与えている。

本報告ではさらに非最適政策の除去可能性を保証した強い結果を与える。以下では便宜的に $\underset{M}{\leqq}$, $\underset{M}{\leqq}$ を $\leqq$, $\leqq$ と記す。

定理2. 任意の $f \in F$ に対し

(i) $S_0(f) \neq \emptyset$ とする。ある $g \in G(f)$ に対して，もし $S_1(f) \cup (S_2(f) \cap C_e(g)) \neq \emptyset$ ならば $u(f) \leqq u(g)$ が成立する。

さらにもし $\hat{S}_0(f) \neq \emptyset$ ならば，任意の $d \in D(f)$ に対して $u(d) \leqq u(f) \leqq u(g)$ が成立する。

(ii) $\tilde{S}_0(f) \neq \emptyset$ とする。各 $d \in D(f)$ に対して，もし $\tilde{S}_1(f) \cup (\tilde{S}_2(f) \cap C_e(d)) \neq \emptyset$ ならば $u(d) \leqq u(f)$ が成立する。さらにもし $S_0(f) \neq \emptyset$ ならば，任意の $g \in G(f)$ に対して $u(d) \leqq u(f) \leqq u(g)$ が成立する。

(証明) 補題4.と5.より容易に示すことができる。

定理2はまた平均型基準値に限定して K-最適性の判定を行うことを可能にしている。相対値の優劣を直接的に評価しない点は従来の定理1による場合と異なる特長的な点である。

§5. 最適方程式とK－最適政策

本節ではK－最適政策が最適方程式の最大解を与える政策として得られることを結論づける。

定義5. YをR^pの任意の有限部分集合とする。凸錐Λに関するYの極大元の全体を集合

$$e[Y|\Lambda] \equiv \{y \in Y \mid ({}^{\forall}x \in Y)(x-y \in \Lambda \rightarrow x=y)\}$$

によって定義する。

次に最適方程式を導入しその最大解を定義する。

定義6. 任意の対$(u,v) \in M^p(S) \times M^p(S)$に対して

$$(OE_1) \qquad u_i \in e\Big[\bigcup_{a \in A_i}\{\sum_{j \in S} q_{ij}^a u_j\} \mid \Lambda\Big],\ i \in S$$

$$(OE_2) \qquad u_i + v_i \in e\Big[\bigcup_{a \in E(i)}\{r_i^a + \sum_{j \in S} q_{ij}^a v_j\} \mid \Lambda\Big],\ i \in S$$

ただし

$$E(i) \equiv \{a \in A_i \mid \sum_{j \in S} q_{ij}^a u_j = u_i\},\ i \in S$$

とする。(OE_1),(OE_2)を最適方程式と呼ぶ。

定義7. 対$(u,v) \in M^p(S) \times M^p(S)$が方程式$(OE_1)$,$(OE_2)$を満足するとき，$(u,v)$を最適方程式の解という。

定理3. 各$f \in F$に対し，対$(u(f),v(f)) \in M^p(S) \times M^p(S)$を求める。対は (i) もし$S_0(f)=\emptyset$ならば，最適方程式の解である，

(iii) もし最適方程式の解でないならば，$S_0(f) \neq \emptyset$である。

(証明) 参考文献[5]と同様に示すことができる。

定義8. 対$(u^*, v^*) \in M^P(S) \times M^P(S)$が最大解であるとは

(i) 最適方程式$(OE_1), (OE_2)$の解である,

(ii) 任意に他の解(u, v)が与えられたとき, uに関して

$(\forall i \in S)(u_i^* \leqq u_i \rightarrow u_i \leqq u_i^*)$を満足する

ことであるとする。

定理4. もしある$f^* \in F$に対して, 対$(u(f^*), v(f^*)) \in M^P(S) \times M^P(S)$が最大解であるならば, f^*はK-最適政策である。

(証明)任意に$f \in F$をとり$(u(f), v(f))$を求める。任意に$i_0 \in S$をとり$u(f^*)_{i_0} \leqq u(f)_{i_0}$とする。もし$(u(f), v(f))$が解ならば$(u(f^*), v(f^*))$が最大解であることにより$u(f)_{i_0} \leqq u(f^*)_{i_0}$が成立する。他方もし$(u(f), v(f))$が解でないならば定理3の(ii)により, $S_0(f) \neq \emptyset$である。さらに定理1により$(u(f), v(f)) \leqq (u(g_1), v(g_1)) \leqq \cdots \leqq (u(g_m), v(g_m))$かつ$S_0(g_m) = \emptyset$とできる。直ちに$u(f^*)_{i_0} \leqq u(f)_{i_0} \leqq u(g_m)_{i_0}$を得る。また$S_0(g_m) = \emptyset$と定理3の(i)より$(u(g_m), v(g_m))$は最適方程式の解である。再び$(u(f^*), v(f^*))$が最大解であることにより$u(g_m)_{i_0} \leqq u(f^*)_{i_0}$である。ゆえに$u(f)_{i_0} \leqq u(f^*)_{i_0}$が成立する。$i_0$は任意より, 任意の$i \in S$について$u(f^*)_i \leqq u(f)_i$ならば$u(f)_i \leqq u(f^*)_i$が成立する。ゆえに$u(f^*)_i - u(f)_i \in K \cap (-K)$よって$u(f^*)_i \simeq u(f)_i$が全ての$i$について成立するので$f^*$はK-最適政策である。

§6. アルゴリズム

本節ではパレート最適政策の全体から成る集合を求めるための非最適政策の除去を伴うアルゴリズムを組み立てる。

K－最適政策の集合を第1近似解集合としてまず求め，これを基礎にパレート最適政策の集合を定める。また定理2による非最適政策の除去法を組み込む。

K_1, K_2 を R^p における任意の凸錐とし，$K_1 \cap (-K_1) = \{0\}$，$K_2 \cap (-K_2) = \{0\}$ なる条件を満足するとする。このとき次の補題を得る。その系と定理1.，2. はアルゴリズムを構成する。

補題6. (P.L. Yu [9])

もし $K_1 \subset K_2$ ならば，$e[Y|K_1] \subset e[Y|K_2]$ である。

系6.1. K_1 およびKによって定義された Λ に関して

もし $R_+^p \subset K_1 \subset \Lambda$ であるならば $e[Y|\Lambda] \subset e[Y|K_1] \subset e[Y|R_+^p]$ が成立する。

この系は，Kを，点0を境界点に持つ閉半空間とし $R_+^p \subset \Lambda$ を満足するように選べば，$e[Y|\Lambda]$ は $e[Y|R_+^p]$ を求めるための近似解集合としうることを示している。ゆえにスカラー化最適性をパレート最適性および K_1－最適性の補助手段として用いることが可能である。$p=2$ のとき，$\{x=(x_1, x_2) \in R^2 \mid x_1 + x_2 \geq 0\}$ はKとして採用しうるが，$K' \equiv \{x=(x_1, x_2) \in R^2 \mid x_2 \geq 0\}$ に対しては $\{K' \cap (-K')^c\} \cup \{0\} \not\supset R_+^2$ より採用しえない。

Algorithm

Step 0. $F_u := F$, $F_b := \emptyset$, $F_\ell := \emptyset$, $F_e := \emptyset$, $F_m := \emptyset$ とおく。

ただし F_u は未調査政策の全体，F_b は doubtful policy の全体，F_ℓ は除去ずみ政策の全体，F_e は等基準値政策の全体，F_m は最適方程式の解を与える政策の全体 を表わす集合とする。 $R_+^p \subset \Lambda$ なる K を任意に与える。

Step 1. もし $F_b = \emptyset$ ならば Step 2 へ；$F_b \neq \emptyset$ ならば g を F_b の中の任意の政策にとり Step 3 へ.

Step 2. もし $F_u \neq \emptyset$ ならば g を F_u の中の任意の政策にとり Step 3 へ；$F_u = \emptyset$ ならば Step 14 へ.

Step 3. $n := 0$, $f_0 := g$, $W_b(g) := \emptyset$, $W_\ell(g) := \emptyset$, $W_e(g) := \emptyset$ とおく。ただし，集合 $W_b(g)$, $W_\ell(g)$, $W_e(g)$ は g 以後の調査により発見された以下の意味を持つ政策の全体を表わす。

$W_b(g)$ は doubtful policy の全体。$W_\ell(g)$ は除去ずみ政策の全体。$W_e(g)$ は等基準値政策の全体。

Step 4. 各 n に対し $f := f_n$ とおき，$(u(f), v(f))$ を求め，$S_0(f)$, $\hat{S}_0(f)$, $G(f)$, $D(f)$ を作る。もし $S_0(f) \neq \emptyset$ ならば $G(f) \neq \emptyset$ より Step 5 へ；$S_0(f) = \emptyset$ ならば $G(f) = \emptyset$ より Step 8 へ.

Step 5. $G(f)$ の中の任意の政策 g をとる。もし $Q(g)u(f) \geq u(f)$ ならば $u(g) \geq u(f)$ より Step 6 へ；$Q(g)u(f) = u(f)$ ならば $L(g, f) \geq 0$ より Step 7 へ.

Step 6. f および $D(f)$ の全ての政策は除去してよいので
$W_\ell(g) := W_\ell(g) \cup D(f) \cup We(g) \cup \{f\}$, $We(g) := \emptyset$ とし
Step 12 へ.

Step 7. $L(g,f)_i \geq 0$ なる i が少くとも 1 つ $Ce(g)$ の中に存在するならば $u(g) \geq u(f)$ であるので Step 6 へ; $L(g,f)_i \geq 0$ なる i は全て $C_o(g)$ に属するならば $u(g) = u(f)$ となり, f は除去できないので $We(g) := We(g) \cup \{f\}$ とする。

Step 8. もし $D(f) \neq \emptyset$ ならば Step 9 へ; $D(f) = \emptyset$ ならば Step 12 へ.

Step 9. $D(f)$ の中の任意の政策をとり d とし, $D(f) := D(f) - \{d\}$ とおく。もし $Q(d)u(f) \leq u(f)$ ならば $u(d) \leq u(f)$ であるので Step 10 へ; $Q(d)u(f) = u(f)$ ならば $L(d,f) \leq 0$ であるので Step 11 へ.

Step 10. d は除去できるので $W_\ell(g) := W_\ell(g) \cup \{d\}$ とし
Step 8 へ.

Step 11. $L(d,f)_i \leq 0$ なる i が少くとも 1 つ $Ce(d)$ の中に存在するならば $u(d) \leq u(f)$ であるので Step 10 へ;
$L(d,f)_i \leq 0$ なる i は全て $C_o(d)$ に属するならば $u(d) = u(f)$ となり d は除去できないので $We(g) := We(g) \cup \{d\}$ とし Step 8 へ.

Step 12. もし $S_o(f) = \emptyset$ ならば Step 13 へ; $S_o(f) \neq \emptyset$ ならば
$f_{n+1} := g$, $n := n+1$ Step 4 へ.

Step13. $B(f)$を作る。$F_m := F_m \cup \{f\}$, $F_e := F_e \cup W_e(\varphi)$,
$W_b(\varphi) := W_b(\varphi) \cup B(f) - (F_m \cup F_e \cup F_\ell)$,
$F_b := F_b - (F_m \cup F_e \cup F_\ell)$, $F_u := F_u - (F_m \cup F_e \cup F_\ell)$
とおく。もし$F_b \neq \emptyset$ならば$n := 0$とおき，f_nをF_bの中の任意の政策にとりStep4へ；$F_b = \emptyset$ならば，
$F_b := W_b(\varphi)$, $F_\ell := F_\ell \cup W_\ell(\varphi)$ とおきStep1へ.

Step14. $F_m^* := \{f \in F_m \mid (u(f), v(f))$は最大解$\}$
$F_e^* := \bigcup_{f \in F_m^*} \{h \in F_e \mid u(h) = u(f)\}$, $F^* := F_m^* \cup F_e^*$
とおく。

F^*はK－最適政策の全体から成る集合である。パレート最適政策の全体を求める必要があればStep15へ；その他はStep16へ.

Step15. $K := R_+^p$とおき $\Lambda_0, \Lambda_1, \Lambda$ を求める。
$F_b := F_m^*$, $F_u := F - (F_m \cup F_e)$, $F_\ell := \emptyset$とおきStep1へ.

Step16. 終了。

数値計算例は口頭発表にて示したので省略する。

謝辞 本報告は九州大学の古川長太教授の御指導の下に行っている研究の一部である。根幹部での御指導のあった事を記し，心より深く感謝の意を表します。

References

[1] D.Blackwell, Discrete Dynamic Programming. Ann.Math.Statis. 33,p.719-726,(1962).

[2] C.H.Cooms,R.M.Daws and A.Tversky, Mathematical Psychology-An Elementary Introduction-. (1970),Prentice-Hall. Translated in Japanese by S.Ono,(1973).

[3] N.Furukawa, Vector-Valued Markovian Decision Processes with Countable Satate Space. Recent Developement in Markov Decision Processes,p.205-223. (1980),Academic Press.

[4] N.Furukawa, Characterization of optimal policies in vector-valued Markovian decision processes. Math.Oper.Res.5, p.271-279,(1980).

[5] T.Iki and N.Furukawa, Vector-Valued Markov Decision Processes with Average Criterion. Mem.Fac.Education Miyazaki University, Nat.Sci.Vol.54.55,p.1-10,(1984).

[6] L.C.Thomas, Constrained Markov decision processes as multi-objective. Multi-Objective Decision Making,p.77-94,(1983).

[7] A.F.Veinott,Jr., On finding optimal policies in discrete dynamic programming with no discounting. Ann.Math.Statis. 37,p.1284-1294,(1969).

[8] YAU-CHUEN WONG and KUNG-FU NG, Partially Ordered Topological Vector Spaces. (1973),Clarendon Press. Oxford.

[9] P.L.Yu, Cone Convexity, Cone Extreme Points, and Nondominated Solutions in Decision Problems with Multiobjectives. J.Opt.Theor.Appl,Vol.14-3,p.319-376, (1974).